



TITLE:

核関数のフーリエ級数展開による U-,V-統計量の解析について(漸近的 統計理論)

AUTHOR(S):

金川, 秀也

CITATION:

金川, 秀也. 核関数のフーリエ級数展開によるU-,V-統計量の解析について(漸近的統計理論). 数理解析研究所講究録 2003, 1308: 185-198

ISSUE DATE:

2003-02

URL:

<http://hdl.handle.net/2433/42859>

RIGHT:

核関数のフーリエ級数展開による U-, V-統計量の解析について

武蔵工業大学・工学部 金川 秀也 (Shuya Kanagawa)

Faculty of Engineering,
Musashi Institute of Technology

§ 0. 序

$\{\xi_j, j \geq 1\}$ を分布 μ に従う実数値確率変数列とする. また $u(x_1, x_2, \dots, x_k)$ を実対称関数とする. $u(x_1, x_2, \dots, x_k)$ を核関数とする統計量を対称統計量と呼ぶ. 対称統計量の一種である U-統計量あるいは V-統計量の漸近的な性質を核関数のフーリエ級数展開によって調べる方法について解説する.

Example 1. U-統計量 (degree k) : $U_n := \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} u(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_k})$

Sample mean : $u(x) = x \quad (k=1)$

Sample variance : $u(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2 \quad (k=2)$

Example 2. V-統計量 (degree k) : $V_n := n^{-k} \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n} u(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_k})$

Cramér-von Mises-Smirnov statistic :

$$u(x_1, x_2) = \int_0^1 w(u) (I_{\{x_1 \leq u\}} - u) (I_{\{x_2 \leq u\}} - u) du \quad (k=2)$$

このとき

$$V_n = \frac{1}{n-1} \int_0^1 w(x) (F_n(x) - 1)^2 dx,$$

ただし $w(x)$ は重み関数、 $F_n(x)$ は $\{\xi_j, 1 \leq j \leq n\}$ の経験分布関数.

任意の x_2, \dots, x_k に対して

$$\int_{-\infty}^{\infty} u(y, x_2, \dots, x_k) \mu(dy) = 0$$

のとき $u(x_1, x_2, \dots, x_k)$ は退化しているという。

§ 1. degree 2 の退化した核関数のフーリエ級数展開

$\{\xi_j, j \geq 1\}$ が i.i.d. の場合は Hoeffding(1948) による H-分解によって退化していない $u(x_1, x_2, \dots, x_k)$ に関する U-統計量、V-統計量について中心極限定理（漸近正規性）、Donsker 型不偏原理、概収束型不偏原理、漸近展開、大偏差原理など極めて精密に漸近的な性質が調べられている。また Yoshihara(1976) によって $\{\xi_j, j \geq 1\}$ が ϕ -mixing 性や absolutely regular 性程度の弱従属性を持つ場合は $\{\xi_j, j \geq 1\}$ を i.i.d. 確率変数列で近似することによって H-分解を用いて退化していない $u(x_1, x_2, \dots, x_k)$ に関する U-統計量、V-統計量について中心極限定理が成り立つことが示されている。弱従属確率変数列の独立確率変数列による近似については Yoshihara (1993) を参照のこと。

一方、退化している核関数 $u(x_1, x_2, \dots, x_k)$ に対して $\{\xi_j, j \geq 1\}$ が i.i.d. や弱従属の場合は Dynkin-Mandelbaum (1983)、Dehling (1983)、Kanagawa-Yoshihara (1993)、Dehling-Denker-Philipp (1994)、Kanagawa and Yoshihara (1994)、Kanagawa (1999) によって各種の極限定理が調べられている。 $\{\xi_j, j \geq 1\}$ が退化している場合や強い従属性を持つ場合は直接 H-分解が使えないために何らかの代わりの方法を見つける必要があるが、その一つとして $u(x_1, x_2, \dots, x_k)$ をフーリエ級数展開し、各種の漸近的な性質を調べることにについて考察する。

degree 2 で核関数 $u(x_1, x_2)$ が退化している場合は Serfling(1980) によって次のような固有値と固有関数を用いたフーリエ級数展開が示されている。

定理 1. $u(x_1, x_2)$ は $R \times R$ 上の対称で ξ_j の分布 μ について 2 乗可積分実数値関数で退化しているとする、i.e. 任意の x に対し $E(u(\xi_1, x)) = 0$ 。 $f \in L^2 = L^2(d\mu \times d\mu)$ に対して $T_u f(x) := E[u(\xi_1, x) f(\xi_1)]$ によって $T_u: L^2 \rightarrow L^2$ (trace class) を定義すると、 T_u は固有関数 $\{g_i\}$ と固有値 $\{\lambda_i\}$ を持つ。これらが次の条件を満足すると仮定する。

$$\begin{cases} E(g_i(\xi_1)) = 0, & E(g_i^2(\xi_1)) = 1 \\ E(g_i(\xi_1)g_j(\xi_1)) = 0 \ (i \neq j), & E(h(\xi_1, x)g_i(\xi_1)) = \lambda_i g_i(x) \end{cases}$$

この時 $L^2(d\mu \times d\mu)$ 収束の意味で $\sum_{k=1}^N \lambda_k g_k(x_1)g_k(x_2)$ は $u(x_1, x_2)$ に収束する、

$$\lim_{N \rightarrow \infty} \iint \left(u(x_1, x_2) - \sum_{k=1}^N \lambda_k g_k(x_1) g_k(x_2) \right)^2 d\mu(x_1) d\mu(x_2) = 0. \quad (1)$$

Mercerの定理より $u \in L^2$ が連続な非負定値関数の場合に各点収束の意味で(1)が収束することが知られているが、最近佐藤坦氏によって区分的に連続な関数の場合にも成立することが示された。

定理 2. 佐藤(1992) X は可分距離空間、 μ は X 上のボレル測度、 $u(\cdot, \cdot)$ は $X \times X$ 上の関数で、 $\mu(X \setminus X_0) = 0$ である μ -可測集合 $X_0 \subset X$ に対して、 $X_0 \times X_0$ 上で $u(\cdot, \cdot)$ が連続とする。この時

$$\iint_{X \times X} |u(x, y)|^2 \mu(dx) \mu(dy) < \infty \quad (2)$$

$$\iint_{X \times X} u(x, y) f(x) \overline{f(y)} \mu(dx) \mu(dy) \geq 0, \quad f \in L^2(d\mu \times d\mu) \quad (3)$$

このとき積分作用素 $T_u f(x) := E[u(\xi_1, x) f(\xi_1)]$ が nuclear であるための必要十分条件は

$$\int_X u(x, x) \mu(dx) < \infty \quad (4)$$

である。

定理2 (Mercer-佐藤の定理) より μ がルベグ測度について絶対連続である場合は階段関数のような区分的に連続な関数 $u(x, y)$ に対して(1)が各点連続の意味で収束する。

系 1. $\mu(X \setminus X_0) = 0$ である μ -可測集合 $X_0 \subset X$ に対して、 $X_0 \times X_0$ 上で $u(\cdot, \cdot)$ が連続とする。(2), (3), (4) の条件の下で

$$u(x_1, x_2) = \sum_{k=1}^{\infty} \lambda_k g_k(x_1) g_k(x_2) \quad \text{a.e. } \mu(dx) \mu(dy) \quad (5)$$

またヒルベルト空間 H を次のように定義する

$$H := \left\{ \mathbf{x} = (x_1, x_2, \dots) \in \mathbf{R}^{\infty} : \sum_{k=1}^{\infty} |\lambda_k| x_k^2 < \infty \right\}$$

$$\langle x, y \rangle := \sum_{k=1}^{\infty} |\lambda_k| x_k y_k \quad (6)$$

$$\|x\| := \left(\sum_{k=1}^{\infty} |\lambda_k| x_k^2 \right)^{1/2}$$

固有値についての条件

$$\sum_{k=1}^{\infty} |\lambda_k| < \infty \quad (7)$$

の下で

$$E \left(\sum_{k=1}^{\infty} |\lambda_k| g_k^2(\xi_i) \right) = \sum_{k=1}^{\infty} |\lambda_k| E(g_k^2(\xi_i)) = \sum_{k=1}^{\infty} |\lambda_k| < \infty,$$

が成り立つので H -値確率変数 $G_i := (g_1(\xi_i), g_2(\xi_i), g_3(\xi_i), \dots)$ $i \geq 1$ を用いて $\{V_n, n \geq 1\}$ を書き直すことができる.

$$x \in H := \left\{ x = (x_1, x_2, \dots) \in \mathbf{R}^{\infty} : \sum_{k=1}^{\infty} |\lambda_k| x_k^2 < \infty \right\}$$

に対して

$$h(x) := \sum_{k=1}^{\infty} \lambda_k x_k^2$$

とおくと、

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n G_i = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g_1(\xi_i), \frac{1}{\sqrt{n}} \sum_{i=1}^n g_2(\xi_i), \frac{1}{\sqrt{n}} \sum_{i=1}^n g_3(\xi_i), \dots \right) \in H. \quad (8)$$

$$nV_n = \frac{1}{n} \sum_{1 \leq i, j \leq n} u(\xi_i, \xi_j) = \frac{1}{n} \sum_{1 \leq i, j \leq n} \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j) \quad (9)$$

$$= \frac{1}{n} \sum_{k=1}^{\infty} \lambda_k \sum_{1 \leq i, j \leq n} g_k(\xi_i) g_k(\xi_j) = \frac{1}{n} \sum_{k=1}^{\infty} \lambda_k \left\{ \sum_{i=1}^n g_k(\xi_i) \right\}^2 = h \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n G_i \right).$$

ここで H -値確率変数 $G_i := (g_1(\xi_i), g_2(\xi_i), g_3(\xi_i), \dots)$ に関する極限定理を用いて $\{V_n, n \geq 1\}$ の漸近的な性質を調べることができる.

§ 2. degree 2 の U -統計量、 V -統計量の表現

$U_n := \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} u(\xi_i, \xi_j)$, $V_n := \frac{1}{n^2} \sum_{1 \leq i, j \leq n} u(\xi_i, \xi_j)$ について

$$\begin{aligned}
 nV_n &= \frac{1}{n} \sum_{1 \leq i, j \leq n} u(\xi_i, \xi_j) = \frac{1}{n} \sum_{1 \leq i, j \leq n} \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j) \\
 &= \frac{1}{n} \sum_{k=1}^{\infty} \lambda_k \sum_{1 \leq i, j \leq n} g_k(\xi_i) g_k(\xi_j) = \frac{1}{n} \sum_{k=1}^{\infty} \lambda_k \left\{ \sum_{i=1}^n g_k(\xi_i) \right\}^2 \\
 &= h\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n G_i\right).
 \end{aligned} \tag{10}$$

同様に

$$\begin{aligned}
 \sqrt{n(n-1)}U_n &= \frac{2}{\sqrt{n(n-1)}} \sum_{1 \leq i < j \leq n} u(\xi_i, \xi_j) \\
 &= \frac{2}{\sqrt{n(n-1)}} \sum_{1 \leq i < j \leq n} \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j) \\
 &= \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \sum_{1 \leq i < j \leq n} 2g_k(\xi_i) g_k(\xi_j) \\
 &= \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \left[\left\{ \sum_{i=1}^n g_k(\xi_i) \right\}^2 - \sum_{i=1}^n g_k^2(\xi_i) \right] \\
 &= \frac{n}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \left[\left\{ \sum_{i=1}^n \frac{g_k(\xi_i)}{\sqrt{n}} \right\}^2 - \frac{1}{n} \sum_{i=1}^n \left(\frac{g_k(\xi_i)}{\sqrt{n}} \right)^2 \right] \\
 &= \frac{n}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \left\{ \sum_{i=1}^n \frac{g_k(\xi_i)}{\sqrt{n}} \right\}^2 - \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \sum_{i=1}^n \frac{g_k^2(\xi_i)}{n} \\
 &= \frac{n}{\sqrt{n(n-1)}} h\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n G_i\right) - \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{\infty} \lambda_k \sum_{i=1}^n \frac{g_k^2(\xi_i)}{n}.
 \end{aligned} \tag{11}$$

§ 2. degree k ($k \geq 3$) の退化した核関数のフーリエ級数展開

次にdegreeが3以上の核関数 $u(x_1, x_2, \dots, x_k)$ に関するU-統計量やV-統計量につ

いても、核関数のフーリエ級数が各点収束すれば、degree 2の場合と同様にそれらの統計量が無限次元確率変数列の和を用いて表現できることを示す。まず

$u(x_1, x_2, \dots, x_k)$ のフーリエ級数展開について考察する。以下の議論は、高橋陽一郎「実解析とFourier解析1」、岩波講座、第二章を参考とした。

$T = [0, 1]$ 、 $C(T)$ を T 上の連続関数全体、 $C(T)$ 上の内積を $\langle f, g \rangle = \int_T f(t) \overline{g(t)} dt$, $f, g \in C(T)$ 、 e_1, e_2, \dots を $C(T)$ 上の正規直行系とする。 $u: T^m \rightarrow \mathbf{R}$ に関するフーリエ係数は、 $k = (k_1, k_2, \dots, k_m) \in \mathbf{Z}^m$ (\mathbf{Z} は整数全体) として

$$\hat{u}(k) = \int_{T^m} u(t) \overline{e_k(t)} dt$$

によって与えられる。ただし、 $t = (t_1, t_2, \dots, t_m) \in T^m$ に対して

$$e_k(t) = e_{k_1}(t_1) e_{k_2}(t_2) \cdots e_{k_m}(t_m) = \exp \left(2\pi i \sum_{j=1}^m k_j t_j \right).$$

もし $u \in L^1(T^m)$ ならばフーリエ係数はFubini の定理によって累次積分で表される。

$$S_N(u, t) = \sum_{k_1=-N}^N \cdots \sum_{k_m=-N}^N \hat{u}(k) e_k(t),$$

$$\langle f, g \rangle = \int_{T^m} f(t) \overline{g(t)} dt, \quad f, g \in C(T^m),$$

$$\|f\|_2 = (\langle f, f \rangle)^{1/2},$$

とおくと $u \in L^1(T^m)$ のフーリエ級数展開は次のように表される。

定理 3. $u \in C(T^m)$ のとき

$$\min_{\{c_k\}} \left\| u - \sum_{k_1=-N}^N \cdots \sum_{k_m=-N}^N c_k e_k \right\|_2^2 = \|u\|_2^2 - \sum_{k_1=-N}^N \cdots \sum_{k_m=-N}^N |\hat{u}(k)|^2$$

$$\|u\|_2 = \left(\sum_{k \in \mathbf{Z}^m} |\hat{u}(k)|^2 \right)^{1/2}.$$

定理 4. $u \in C^m(T^m)$ のとき、 $t = (t_1, t_2, \dots, t_m) \in T^m$ に関して一様に

$$S_N(u, t) = \sum_{k_1=-N}^N \cdots \sum_{k_m=-N}^N \hat{u}(k) e_k(t) \rightarrow u(t) \quad (N \rightarrow \infty). \quad (12)$$

定理 3 では、 $u \in C^m(T^m)$ に対して各点一様収束が成り立つので従属確率変数列 $\{\xi_j, j \geq 1\}$ に対し、 $\xi = (\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_m})$ と置いたとき、確率 1 で

$$S_N(u, \xi) = \sum_{k_1=-N}^N \cdots \sum_{k_m=-N}^N \hat{u}(k) e_{k_1}(\xi_{i_1}) e_{k_2}(\xi_{i_2}) \cdots e_{k_m}(\xi_{i_m}) \rightarrow u(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_m}) \quad (N \rightarrow \infty)$$

が成り立つ。しかし、Cramér-von Mises-Smirnov statistic では核関数

$$u(x_1, x_2) = \int_0^1 w(u) \left(I_{\{x_1 \leq u\}} - u \right) \left(I_{\{x_2 \leq u\}} - u \right) du \quad (13)$$

が連続（区分的連続）ではないのでこのままでは定理 2 を適用できない。そのため不連続な関数に対するフーリエ級数展開についても調べる必要がある。

以上のように核関数 $u(x_1, x_2, \dots, x_k)$ のフーリエ級数が各点収束すれば、degree 2 の場合と同様に U-統計量や V-統計量が無限次元確率変数列の和を用いて表現できることを示す。

可分なバナッハ空間 X とノルム $\|\cdot\|$ を次の様の定義する。

$$X := \left\{ \mathbf{x} = (x_1, x_2, \dots) \in \mathbb{R}^\infty : \sum_{-\infty \leq k_1, \dots, k_m < \infty} |\hat{u}(k)| |x_{k_1}| \cdots |x_{k_m}| < \infty \right\},$$

$$\|\mathbf{x}\| := \left\{ \sum_{-\infty \leq k_1, \dots, k_m < \infty} |\hat{u}(k)| |x_{k_1}| \cdots |x_{k_m}| \right\}^{1/m}.$$

X 上の連続関数 $\varphi(\mathbf{x})$ を

$$\varphi(\mathbf{x}) := \sum_{-\infty \leq k_1, \dots, k_m < \infty} \hat{u}(k) x_{k_1} \cdots x_{k_m} \quad (14)$$

とおき、 X -値確率変数 $\mathbf{G}_k := (e_1(\xi_k), e_2(\xi_k), \dots)$, $k \geq 1$ とすると、

$$\sum_{k=1}^n \mathbf{G}_k = \left(\sum_{k=1}^n e_1(\xi_k), \sum_{k=1}^n e_2(\xi_k), \dots \right) \quad (15)$$

より、

$$\begin{aligned} n^m V_n &= \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_m \leq n} u(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_m}) \\ &= \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_m=-\infty}^{\infty} \hat{u}(k) \left\{ \sum_{j=1}^n e_{k_1}(\xi_j) \sum_{j=1}^n e_{k_2}(\xi_j) \dots \sum_{j=1}^n e_{k_m}(\xi_j) \right\} \\ &= \varphi \left(\sum_{k=1}^n \mathbf{G}_k \right). \end{aligned} \quad (16)$$

ゆえに $n^m V_n$ が X -値確率変数列 $\{\mathbf{G}_k, k \geq 1\}$ の部分和の汎関数として表すことが出来るので、 $\{\mathbf{G}_k, k \geq 1\}$ に関する極限定理を応用することで V -統計量 V_n の漸近的な性質を調べることが出来る。通常のフーリエ級数展開は固有関数による展開のように核関数 $u(x_1, x_2, \dots, x_k)$ に対称性を仮定する必要がないので、 U -, V -統計量含む、より一般の対称統計量にも面倒な核関数の対称化を行わずに応用できる。

§ 3. 大偏差原理への応用

核関数のフーリエ級数展開にヒルベルト空間値確率変数列に関する極限定理を応用することで U -, V -統計量の漸近的な性質を従来の方法よりも明解に考察することが出来る。Cramér-von Mises-Smirnov 統計量のような退化した不連続な核関数を持つ対称統計量に関する重複対数の法則や大偏差理論のような概収束性を論じる場合に、定理 1 が 2 乗平均の意味での収束であることから確率変数列に対していくつかのテクニカルな条件を仮定する必要があった。しかし定理 2 (Mercer-佐藤の定理) を用いる事によって、そのような条件を省略することが出来る。まず大偏差原理について解説する。

定理 5. (Cramér の定理) $\{\xi_j, j \geq 1\}$ i.i.d. 実数値確率変数列で平均 0, 分散 1 とする。また $E\{\exp(t\xi_1)\} < \infty$, $t > 0$ とする。このとき $S_n = \sum_{i=1}^n \xi_i$ とおくと、任意の閉集合 $F \subset \mathbf{R}$ に対して

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log P\{S_n \in F\} \leq - \inf_{x \in F} I(x), \quad (17)$$

また任意の開集合 $G \subset \mathbf{R}$ に対して

$$\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log P\{S_n \in G\} \geq - \inf_{x \in G} I(x), \quad (18)$$

ただし

$$I(x) := \sup_{\theta} [\theta x - \log M(\theta)], \quad M(\theta) := E\{\exp(\theta \xi_1)\}. \quad (19)$$

定理 6. (Donsker-Varadhan (1976)) B をバナッハ空間で $\|\cdot\|$ をノルムとする.

B -値 i.i.d. 確率変数列 $\{\xi_j, j \geq 1\}$ は平均 0 で $E\{\exp(t\|\xi_1\|)\} < \infty, t > 0$ を仮定する. さらに $\Phi: B \rightarrow \mathbf{R}$ は連続関数で、定数 C, D 及び任意の $x \in B$ に対して

$$\Phi(x) \leq C + D\|x\|. \quad (20)$$

このとき

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E\left\{\exp\left(n\Phi\left(\frac{S_n}{n}\right)\right)\right\} = \sup_{x \in B} [\Phi(x) - h(x)], \quad (21)$$

ただし、 $h(x) := \sup_{\varphi \in B^*} [\varphi(x) - \log M(\varphi)]$, B^* は B の dual で $M(\theta) := E\{\exp(\varphi(\xi_1))\}$.

定理 7. (Bolthausen (1986))

定理 8 の条件をすべて仮定する. また $\Phi(x') - h(x') = \sup_{x \in B} [\Phi(x) - h(x)]$ となる $x' \in B$ がただ一つ定まり、 $\Phi(\cdot)$ は B 上で 3 次 Fréchet 微分可能とする. さらにいくつかの仮定の下で

$$\lim_{n \rightarrow \infty} \exp\{-n(\Phi(x') - h(x'))\} E\left\{\exp\left(n\Phi\left(\frac{S_n}{n}\right)\right)\right\} = \int \exp\left\{\frac{1}{2} D^2 \Phi(x')[y, y]\right\} \gamma(dy),$$

ただし、 $\gamma(dy)$ はあるガウス測度、 D^2 は 2 階の Fréchet 微分.

定理 8. (Kanagawa (1999)) $\{\xi_j, j \geq 1\}$ は i.i.d. で分布を μ とする. また degree 2 の

対称核関数 $u(x, y)$ に対して (2), (3), (7) 及び Cramér 条件

$$E(\exp(t\|G_1\|)) < \infty$$

を仮定する. $\mathbf{x} = (x_1, x_2, \dots) \in H$ に対して $\Phi(\mathbf{x}) := \left| \sum_{k=1}^{\infty} \lambda_k x_k^2 \right|^{1/2}$ とおく. G_1 の分布に関するエントロピー関数 h を

$$h(\mathbf{x}) := \sup_{\varphi \in H^*} (\varphi(\mathbf{x}) - \log M(\varphi)),$$

ただし H^* は H の topological dual で $M(\varphi) := E(\exp(\varphi(G_1)))$. このとき

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E(\exp(n|V_n|^{1/2})) = \sup_{\mathbf{x} \in H} (\Phi(\mathbf{x}) - h(\mathbf{x})).$$

さらに μ が絶対連続であるとき

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E(\exp(n|U_n|^{1/2})) = \sup_{\mathbf{x} \in H} (\Phi(\mathbf{x}) - h(\mathbf{x})).$$

$$\text{証明. } H = \left\{ \mathbf{x} = (x_1, x_2, \dots) \in \mathbf{R}^\infty : \sum_{k=1}^{\infty} |\lambda_k| x_k^2 < \infty \right\},$$

$G_i = (g_1(\xi_i), g_2(\xi_i), g_3(\xi_i), \dots) \in H$ 及び、定理 2 (Mercer-佐藤の定理) によって

$$u(\xi_i, \xi_j) = \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j) \quad \text{a.s.}$$

より

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left\{ \exp \left\{ n |V_n|^{1/2} \right\} \right\} &= \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left\{ \exp \left(\left| \sum_{1 \leq i, j \leq n} u(\xi_i, \xi_j) \right|^{1/2} \right) \right\} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left\{ \exp \left(\left| \sum_{1 \leq i, j \leq n} \sum_{k=1}^{\infty} \lambda_k g_k(\xi_i) g_k(\xi_j) \right|^{1/2} \right) \right\} \end{aligned}$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left\{ \exp \left(\left| \sum_{k=1}^{\infty} \lambda_k \sum_{1 \leq i, j \leq n} g_k(\xi_i) g_k(\xi_j) \right|^{1/2} \right) \right\} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left\{ \exp \left(\left| \sum_{k=1}^{\infty} \lambda_k \left(\sum_{i=1}^n g_k(\xi_i) \right)^2 \right|^{1/2} \right) \right\} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left\{ \exp \left(n \left| \sum_{k=1}^{\infty} \lambda_k \left(\frac{1}{n} \sum_{i=1}^n g_k(\xi_i) \right)^2 \right|^{1/2} \right) \right\} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left\{ \exp \left(n \Phi \left(\frac{S_n}{n} \right) \right) \right\},
\end{aligned}$$

ただし $S_n := \sum_{i=1}^n G_i = \sum_{i=1}^n (g_1(\xi_i), g_2(\xi_i), g_3(\xi_i), \dots)$. 一方

$$\Phi(\mathbf{x}) = \left| \sum_{k=1}^{\infty} \lambda_k \mathbf{x}_k^2 \right|^{1/2} \leq \left(\sum_{k=1}^{\infty} |\lambda_k| x_k^2 \right)^{1/2} = \|\mathbf{x}\|$$

より(20)が成り立つため定理8 (Donsker-Varadhan大偏差原理)を用いれば定理が証明される.

系2. μ を $[0,1]$ 上の一様分布とする. Cramér-von Mises-Smirnov統計量の核関数

$$u(x_1, x_2) = \int_0^1 w(u) \left(I_{\{x_1 \leq u\}} - u \right) \left(I_{\{x_2 \leq u\}} - u \right) du$$

に対して定理10が成り立つ.

重み関数 $w(x) = \frac{1}{x(x-1)}$ に対してCramér-von Mises-Smirnov統計量の固有値は

$$\lambda_k = \frac{1}{k(k+1)}, \quad k \geq 1$$

で与えられる (Csáki(1980), Borovskikh(1985)). ゆえに

$$V_n = \frac{1}{n-1} \int_0^1 w(x) (F_n(x) - 1)^2 dx$$

に関するDonsker-Varadhan大偏差原理のエントロピー関数 $h(\mathbf{x})$ は

$$H = \left\{ \mathbf{x} = (x_1, x_2, \dots) \in \mathbf{R}^\infty : \sum_{k=1}^{\infty} \frac{x_k^2}{k(k+1)} < \infty \right\}$$

の topological dual H^* に対して

$$h(\mathbf{x}) = \sup_{\varphi \in H^*} (\varphi(\mathbf{x}) - \log M(\varphi))$$

によって与えられる.

同様に退化した核関数に対するU-統計量についてBolthausenの大偏差原理も成り立つ.

定理 9. (Kanagawa (1999)) 定理 8 と同じ条件を仮定する. また $\Phi(\mathbf{x}') - h(\mathbf{x}') = \sup_{\mathbf{x} \in B} [\Phi(\mathbf{x}) - h(\mathbf{x})]$ となる $\mathbf{x}' \in H$ がただ一つ定まり、 $\Phi(\cdot)$ は H 上で 3 次Fréchet微分可能とする. さらにいくつかの仮定の下で

$$\lim_{n \rightarrow \infty} \exp\{-n(\Phi(\mathbf{x}') - h(\mathbf{x}'))\} E\left\{\exp(n|U_n|^2)\right\} = \int \exp\left\{\frac{1}{2} D^2 \Phi(\mathbf{x}')[y, y]\right\} \gamma(dy),$$

ただし、 $\gamma(dy)$ はあるガウス測度、 D^2 は 2 階のFréchet微分. この定理は V_n についても全く同様に成り立つ.

その他の退化した対称核関数についてのU、V-統計量に関する極限定理について[4]~[7], [9]~[11]を参照されたい.

REFERENCES

- [1] Yu. V. Borovskikh, Estimates of characteristic functions of some random variables with applications to ω^2 -statistics. II, Theory of Prob. Appl., 30 (1985) 117-127.
- [2] E. Csáki, On the standardized empirical distribution function, in B.V. Gnedenko et al. eds., Non-parametric Statistical Inference, Coll. Mathem. Soc. János Bolyai 32 (North-Holland, Amsterdam, 1980) pp. 123-138.
- [3] H. Dehling, Limit theorems for sums of weakly dependent Banach space valued

- random variables, *Wahrscheinlichkeitstheorie verw. Gebiete*, **63** (1983), 393-432.
- [4] H. Dehling, M. Denker and W. Philipp, Invariance principles for von Mises and U-statistics, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **67** (1984), 139-167.
- [5] E. B. Dynkin and A. Mandelbaum, Symmetric statistics, Poisson point processes and multiple Wiener integrals, *Ann. Statist.*, **11** (1983) 739-745.
- [6] F. Götze, On Edgeworth expansions in Banach spaces, *Ann. Probability*, **9** (1981) 852-859.
- [7] M. Harel and M. L. Puri, Limiting behaviour of U-statistics, V-statistics, and one sample rank order statistics for non stationary absolutely regular processes, *J. Multivariate Anal.*, **30** (1989), 181-204.
- [8] W. Hoeffding, A non-parametric test of independence, *Ann. Math. Statist.*, **19** (1948), 546-557.
- [9] S. Kanagawa and K. Yoshihara; The almost sure invariance principles of degenerate U-statistics of degree two for stationary random variables, *Stochastic Processes and their Applications*, **49** (1994), pp347 - pp356
- [10] S. Kanagawa and K. Yoshihara, Almost sure invariance principle for V- and U-statistics based on weakly dependent random variables, *Asymptotic Statistics (Proceedings of the Fifth Prague Symposium, Prague, Czech Republic, 1993)*, Physica-Verlag, (1994) 318 - 327.
- [11] S. Kanagawa, A representation of the rate functions in large deviation principles for U-statistics with degenerate kernels, *Trends in Probability and Related Analysis (Proceedings of Symposium on Analysis and Probability 1998, Tiwan University, Taipei)*, 254-264, (1999).
- [12] H. Sato, Nuclearity of a nonnegative definite integral kernel on a separable metric space, *J. Theoretical Probab.*, **5** (1992), 349-353.
- [13] V. V. Sazonov and V. V. Ul'yanov, Asymptotic expansions of the probability that the sum of independent random variables hits a ball in a Hilbert space, *Russian Math. Surveys*, **50:5** (1995) 1045-1063.
- [14] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, New York, 1980.
- [15] K. Yoshihara, Limiting behaviour of U-statistics for stationary, absolutely regular

processes, Z. Wahrscheinlichkeitstheorie und verw. Gebiete, 35 (1976), 237-252.

- [16] K. Yoshihara, *Weakly Dependent Stochastic Sequences and Their Applications*, Vol. 2, *Asymptotic Statistics based on Weakly Dependent Date* (Sanseido Co. Ltd., Tokyo, 1993). K. Yoshihara, *Weakly Dependent Stochastic Sequences and Their Applications*, Vol. 2, *Asymptotic Statistics based on Weakly Dependent Date* (Sanseido Co. Ltd., Tokyo, 1993).
- [17] 高橋陽一郎、実解析とFourier解析 1、岩波講座、第二章